

Causal Inference 1 (Nonexperimental data)

29 June - 1 July 2020 (12 hours)

Bruno Arpino

Department of Statistics, Computer Science, Applications
University of Florence, Italy

Short biography of the instructor

Bruno Arpino is an associate professor at the Department of Statistics, Computer Science, Applications of the University of Florence (Italy). Previously, he was associate professor at the Department of Political and Social Sciences, Universitat Pompeu Fabra (UPF) and co-director of the Research and Expertise Centre on Survey Methodology (RECSM, UPF). He obtained a PhD in Applied Statistics from the University of Florence in 2008. His main research interests are in the areas of causal inference, applied statistics and social demography. He is particularly interested in propensity score methods, especially for multilevel and network data. From a substantive point of view, he has been studying intergenerational relationships, ageing and health, fertility and immigrants' assimilation. He has published articles in international peer-reviewed journals such as *The Annals of Applied Statistics*, *Demography*, *European Sociological Review*, *Journal of Marriage and Family*, *Journal of the Royal Statistical Society - A and C*, *Social Indicators Research*, *Statistics in Medicine*. Since October 2017 he is member of the Editorial Board of *Statistical Methods and Applications*.

Web-page: <https://sites.google.com/site/brunoarpino/>

Course description

What is the effect of smoking on health? Does having an additional child increase the risk of poverty? Are development policies targeted on small firms effective in increasing investments?

Most studies in the social sciences are motivated by questions that are causal in nature.

In these areas, however, experiments are not always possible because of ethical or practical reasons and the estimation of causal effects has often to rely on observational studies. The validity of inference will then strictly depend on the plausibility of the assumptions underlying the employed statistical techniques.

This course will cover some of the most popular techniques for estimating causal effects with observational data: propensity score matching, instrumental variable regression, regression discontinuity designs and fixed effects models. Special emphasis will be placed during the course on discussing the plausibility of the identifying assumptions,

the data requirements and other practical and theoretical challenges for the implementation of each method.

This short course will offer participants theoretical and applied perspectives on the covered topics. Examples will be drawn from political science, sociology, economics, public health and policy evaluation. Lab sessions will demonstrate the implementation of the covered techniques using the software STATA.

Software

STATA will be used in the lab sessions. Although previous experience of working in STATA is preferable, this is not strictly necessary. Easy to follow do-files and dataset to replicate the analyses discussed in class will be provided.

Prerequisites

Familiarity with regression analysis and basic statistical concepts.

Schedule

June 29, 2020

- Why do we need causal inference?
 - Introduction and course overview
- What assumptions do we impose to identify causal effects?
 - Potential Outcome framework
 - Good Causal Inference is Good Theory + Data + Methods
 - Randomized experiments versus observational studies
- What methods can we use to estimate causal effects?
 - Overview of statistical methods corresponding to different sets of assumptions
- Brief recap on regression methods: pros and cons
- Propensity score methods (matching and weighting):
 - Assumptions
 - Implementation
 - Examples from published research
 - Pros and cons
- Let's put our hands on!
 - Guided lab on propensity score matching
- Summary and next class "trailer"

June 30, 2020

- Brief resume of previous class
- What is and how to implement a sensitivity analysis
- Instrumental Variable Regression:
 - Assumptions
 - Implementation
 - Examples from published research
 - Pros and cons
- Let's put our hands on!
 - Guided lab on Instrumental Variable Regression

- Summary and next class “trailer”

July 1, 2020

- Brief resume of previous class
- Fixed effects models
 - Assumptions
 - Implementation
 - Examples from published research
 - Pros and cons
- Let’s put our hands on!
 - Guided lab on Fixed Effects Models
- Discussion on: Nonstandard data structures and more advanced topics
 - What variables should be used as control variables?
 - Matching as a pre-processing tool
 - Matching with clustered, network or longitudinal data
 - Instrumental Variable analysis without (good) instrument
- Summary
 - Key take home messages
 - Indications for current and future developments in causal inference for the social sciences (causal inference and machine learning)

References:

Becker, S.O. (2016). Using instrumental variables to establish causality. IZA World of Labor.

Cameron, A., & Trivedi, P. (2010). Microeconometrics using Stata. College Station: Stata Press.

Guo, S., and Fraser, M.S. (2009). Propensity score analysis: Statistical methods and applications. Sage Publications.

Imbens, G.W. & Rubin, D.B. (2015). Causal inference for statistics, social, and biomedical sciences: An introduction. Cambridge University Press.

Shahidur R. Khandker, Gayatri B. Koolwal and Hussain A. Samad (2010) Handbook on Impact Evaluation. Quantitative Methods and Practices, The World Bank, Washington D.C. Freely available on-line at:

http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2009/12/10/000333037_20091210014322/Rendered/PDF/520990PUB0EPI1101Official0Use0Only1.pdf